KLAST: a new high-performance sequence similarity search tool

Erwan Drezen, Patrick Durand, Dominique Lavenier

KLAST

High-performance sequence similarity search tool

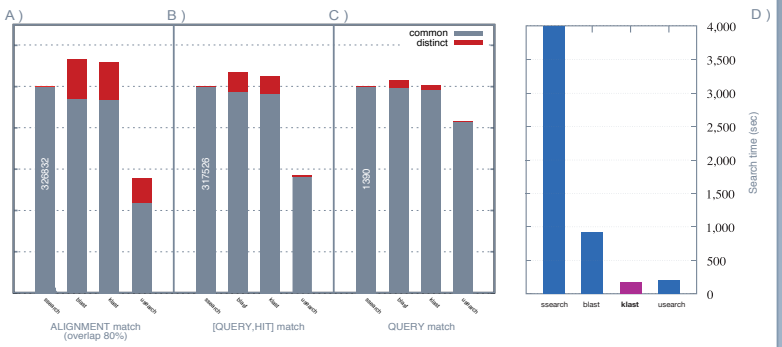
E. DREZEN – P. DURAND – D. LAVENIER

Bio-IT – Boston – April 29th – 1st May 2014

KLAST is a fast, accurate and NGS scalable bank-to-bank sequence similarity search tool providing significant accelerations of seeds-based heuristic comparison methods, such as the Blast suite of algorithms. Relying on unique software architecture, KLAST takes full advantage of recent multi-core personal computers without requiring any additional hardware devices.

- Professional version of PLAST (BMC Bioinformatics, 2009)
- Optimized for bank-to-bank sequence comparisons
- Provide high speed and high quality results

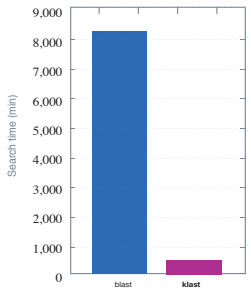
Tools benchmark



The SSEARCH, BLAST, USEARCH software are considered

Comparison of 2,329 protein sequences from bacterium *A.hospitalis* against the SwissProt databank
The reference is given by SSEARCH since it implements the rigorous Smith and Waterman algorithm and generates optimal alignments. Alignments are evaluated on a moderate-size dataset due to the long execution time of SSEARCH. The diagrams synthetize the numbers of alignments found by the different softwares (A, B), the number of queries matching the reference databank (C) and the search time (D).
The gray section identifies similar alignments, i.e. alignments having an overlap greater than 80%. The red section reports other alignments. As it can be seen, the total number of alignments can exceed the number of alignments found by SSEARCH. This is mainly due to the fragmentation of long alignments (including large gaps) into shorter ones by KLAST, BLAST and USEARCH.
More benchmarks are available at koriscale.inria.fr

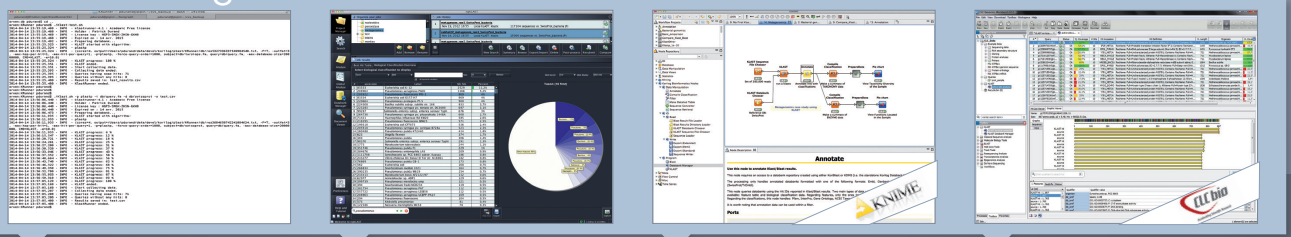
Tara Oceans benchmark



Application on comparative bacterial genomics

KLAST and BLAST benchmark comparison of 8,245 sequences (translated 454 reads) from Tara Oceans metagenomic data against 15 million proteins from Uniprot. Both algorithms ran on 8 Intel Xeon cores. KLAST achieved sequence comparisons 18x times faster than BLAST, while covering up to 96% of the results produced by BLAST.
Benchmark data courtesy of Thomas Vannier and Jean-Marc Aury research team (Genoscope/CEA).
More on this study is available at tinyurl.com/d54ahrb

KLAST integration



Command-line tool | Graphical platform | Workflow and data analytics platform | CLC Genomics workbench

KoriLog
BIOINFORMATICS Solutions
4 rue Gustave Eiffel
56230 Questembert
France
Phone : + 33 960 368 038
www.korilog.com - klast@korilog.com

KLAST software development by

Inria INVENTORS FOR THE DIGITAL WORLD | Région Bretagne | CRITT Santé Bretagne